

HEALTH ABC OUTCOMES DATA ANALYSIS FILE	2
HEALTH ABC EVENTS	3
HOSPITALIZATION DATASET (HOSP.SAS7BDAT)	4
1. General description.....	4
2. Cross reference of dataset names with exact source	6
3. Dataset structure and contents	6
4. Condition of data	6
5. Dataset index formulation and key variable mapping.....	6
6. General strategies for manipulating and merging the data	7
HOSPITALIZATION DETAILS DATASET (HOSPDTLS.SAS7BDAT)	7
1. General description.....	7
2. Cross reference of dataset names with exact source	7
3. Dataset structure and contents	7
4. Condition of data	7
5. Dataset index formulation and key variable mapping.....	8
6. General strategies for manipulating and merging the data	8
DISEASE/CONDITION SPECIFIC DATASETS	8
1. General description.....	8
2. Cross reference of dataset names with exact source	11
3. Dataset structure and contents	11
a. BPH dataset	12
b. Cancer dataset.....	12
c. Cardiovascular dataset.	12
d. Carotid artery disease dataset	13
e. Death dataset	13
f. Diabetes dataset	14
g. Fracture dataset	14
h. GI dataset	15
i. Osteoarthritis / joint surgery dataset	16
j. Peripheral arterial disease dataset	17
k. Psychiatric event dataset:.....	17
l. Pulmonary event dataset	18
m. Stroke/TIA event dataset.....	19
n. "Other" diagnosis dataset.....	19
4. Condition of data	22
5. Dataset index formulation and key variable mapping.....	22
6. General strategies for manipulating and merging the data	22
SPECIAL MISSING VALUE CODES	23

HEALTH ABC OUTCOMES DATA ANALYSIS FILE

The following SAS datasets are now posted on the Health ABC website under Datasets & Documentation/Current Datasets. To use the data, please contact the PI at your site. PIs can download the datasets as self-extracting .exe files from the Health ABC website.

SAS Datasets

Hosp.sas7bdat	Hospitalization events dataset
HospDtls.sas7bdat	Hospitalization details dataset
BPH.sas7bdat	BPH outcomes dataset
Cancer.sas7bdat	Cancer outcomes dataset
Carotid.sas7bdat	Carotid artery disease outcomes dataset
CVATIA.sas7bdat	Stroke/TIA outcomes dataset
CVD.sas7bdat	Cardiovascular disease outcomes dataset
Death.sas7bdat	Death outcomes dataset
Diabetes.sas7bdat	Diabetes mellitus outcomes dataset
Fracture.sas7bdat	Fracture outcomes dataset
GI.sas7bdat	Upper and lower gastrointestinal disease outcomes dataset
Lung.sas7bdat	Pulmonary disease outcomes dataset
OASurg.sas7bdat	Osteoarthritis surgery outcomes dataset
Other.sas7bdat	“Other” outcomes dataset
PAD.sas7bdat	Peripheral arterial disease outcomes dataset
Psych.sas7bdat	Psychiatric outcomes dataset
Formats.sc2	SAS Format Library

HEALTH ABC EVENTS

In Health ABC, an event is defined as an illness episode represented by any overnight hospitalization in an acute care hospital, a newly diagnosed cancer (inpatient or outpatient), a fracture (inpatient or outpatient, excluding ribs, chest/sternum, skull/face, fingers, toes, and cervical vertebra or neck), an outpatient revascularization procedure, or death. An event may comprise more than one hospitalization if the participant is directly transferred from one acute care facility to another acute care facility. An event may be associated with a single diagnosis or multiple diagnoses/conditions.

Participants are asked to report any hospitalizations, outpatient cancer, fracture, or angioplasty events; and every 6 months they are asked directed questions to elicit information about events of this type. Often, additional unreported events are uncovered in the process of researching reported events. These additional events are frequently not captured until the death of a participant is being investigated.

The data contained in the outcomes dataset are based on the best possible documentation abstracted from medical records and other support documents. Therefore, some missing data related to the criteria may be noted. Adjudicator discretion was used to override criteria when some information was missing, but there was clear evidence indicating the clinical event did occur. The “Possible” category was used if evidence for the diagnosis or condition existed but confirmatory information was not available.

All reported deaths at both the Memphis and Pittsburgh sites were adjudicated through the end of all data collection for the HABC study on 9/30/14. There are no un-adjudicated deaths. Therefore, censoring HABC data by this death adjudication through date has no effect. Again, the death adjudication through date and the final data collection date are the same, so all HABC data are included. In other words, there are no data to be censored.

However, adjudication for non-fatal events is only complete for events reported through 8/14/12 for Memphis and Pittsburgh. Morbid events reported after that date continued to be collected, but were not adjudicated. Therefore, for many morbidity analyses it would be cleanest to censor events to include only those with discharge dates no later than 8/14/12. A date of last contact (DTLASTCT), which consists of the date of death, the date of the last completed contact (missed visits not included), or the discharge date for the most recent entered event, whichever comes last, is also included for each participant. If the analyst chooses to censor the data at the date through which adjudication was completed, then the censor date for each participant will be the DTLASTCT or the appropriate date above, whichever comes first.

For those wishing to use the complete data, with the caveat that there will be bias introduced by incomplete adjudication, DTLASTCT may be used for the censor date for all participants.

The D&D Outcomes Dataset consists of 16 individual datasets as follows:

- 1) Hospitalization dataset (Hosp.sas7bdat)
- 2) Details about concurrent hospitalization events dataset (HospDtls.sas7bdat)
- 3) Disease/condition-based datasets:
 - a) Benign prostatic hyperplasia (BPH.sas7bdat)
 - b) Cancer (Cancer.sas7bdat)
 - c) Cardiovascular (CVD.sas7bdat)
 - d) Carotid artery disease (Carotid.sas7bdat)
 - e) Death (Death.sas7bdat)
 - f) Diabetes mellitus (Diabetes.sas7bdat)
 - g) Fracture (Fracture.sas7bdat)
 - h) Gastrointestinal (GI.sas7bdat)
 - i) Osteoarthritis / joint surgery for osteoarthritis (OASurg.sas7bdat)
 - j) Peripheral arterial disease (aorta, iliac arteries, or below) (PAD.sas7bdat)
 - k) Psych (Psych.sas7bdat)
 - l) Pulmonary (Lung.sas7bdat)
 - m) Stroke (CVA) and transient ischemic attack (CVATIA.sas7bdat)
 - n) Other diagnoses (Other.sas7bdat)

Most events, except some cancers, fractures, outpatient revascularizations, and out-of-hospital deaths, are recorded on a Local Adjudication Report, which focuses on details of the hospitalization, as well as criteria for adjudication of an event as one of the Health ABC endpoints. For some events (cancer, cardiovascular events, death, and fracture), the more detailed information is recorded on special adjudication reports specific to that type of event (see annotated forms).

HOSPITALIZATION DATASET (HOSP.SAS7BDAT)

1. General description

The Hospitalization dataset consists of information from the first two pages of the Local Adjudication Report. For each hospitalization the participant has experienced, the following per-hospitalization variables are included (Hx indicates the xth hospitalization, e.g., H1 for the first hospitalization):

HxEVREF	Event Form Reference Number for xth hospitalization
HxADMDT	Admission date for xth hospitalization
HxDISDT	Discharge date for xth hospitalization
HxDXPRM	Primary reason for xth hospitalization
HxLOSTA	Length of stay for xth hospitalization
HxCNCUR	Number of concurrent diagnoses reported during xth hospitalization (includes both “yes” and “possible” events, see page 2 of annotated Local Adjudication Report)
HxDS2EV	Number of days between enrollment (Baseline <u>clinic</u> visit) and xth hospitalization

Please link to the Health ABC participant history (PH) file, available for download on the Health ABC website, for variables such as clinic site, date of birth, gender and race. For each visit/contact there is also a visit date, a variable indicating the type of visit, age at visit and vital status at visit. The vital status variables in the PH file are defined as follows: If a participant had a contact at xx months, or if they missed a contact but they were determined to still be alive at the time their contact was due (participant refused the contact, was too ill, withdrew at contact, etc), then VITALxxM is alive. If they missed a contact and were later discovered to have died before the end of their contact window, then VITALxxM is dead. Finally, if they missed a contact without a determination of their vital status (participant could not be located, withdrew at prior contact, etc) and no further contact with vital status determination has been made since then, then VITALxxM is missing (.Z). If a participant has not yet had the xxM visit, VITALxxM is also .Z. In addition to the visit-specific vital status variables, a “current vital status” variable (that is, vital status as of the creation date of the PH file date) is provided. This PH variable is called VSTATUS.

Date of last contact is the date of either the last clinic, home or proxy visit or phone interview, date of death, or discharge date of last adjudicated event, whichever comes last.

2. Cross reference of dataset names with exact source

A complete list of variable names can be found in Appendix V: Proc Contents (search under HOSP). Variable names can also be found in the annotated forms or in the listing of calculated variables.

3. Dataset structure and contents

The HOSP file contains a single observation per participant and all participants enrolled are represented (n=3075), regardless of whether they have experienced a hospitalization.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
HxEVREF	Event Form Reference Number for the xth hospitalization; this number can be used to identify this event within all the other datasets, particularly the Hospitalization Details dataset (HospDtls.sas7bdat)

4. Condition of data

a. Known data errors: None at this time. The data have been edited.

b. Strengths and weaknesses of dataset items: This dataset is intended primarily for analyses focusing on hospitalization *per se* as an event, rather than for analyses focusing on particular types of events like fractures, cancer, etc. The time to first hospitalization can be determined from H1DS2EV. The number of hospitalizations that had occurred before the second (12-mo/Year 2) clinic visit, for example, can be determined by comparing HxADMDT to CV2DATE. A breakdown of the primary reasons for hospitalization can be made from HxDXPRM, and so on.

What is lost in this format is the ability to look at particular combinations of diagnoses that may have occurred at the same time. For example, the analyst may wish to focus on a particular combination of concurrent diagnoses such as MI complicated by pneumonia. In addition, the analyst may want to

make use of the ICD-9 discharge or diagnosis codes, when available. For that reason, a second dataset (HospDtls.sas7bdat) has been provided, which includes one record per hospital stay (i.e., multiple records per participant). See documentation below for more details about that dataset.

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied.

5. Dataset index formulation and key variable mapping

The HOSP file is sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

HOSPITALIZATION DETAILS DATASET (HOSPDTLS.SAS7BDAT)

1. General description

The Hospitalization details dataset also consists of information from the first two pages of the Local Adjudication Report (see annotated forms). However, this dataset is organized in a one-record-per-event format (i.e., multiple records possible for each participant). For some events, there are two records, see Dataset structure and contents below. In addition, ICD-9 codes from the Discharge Abstract Form have been merged onto each record, when available. If no ICD-9 codes were available, or if the Event Form Reference Number (EVREF) was omitted or incorrectly transcribed on either form (edits are in place to check for this), then the ICD-9 variables will be missing.

2. Cross reference of dataset names with exact source

A complete list of variable names can be found in Appendix V: Proc Contents (search under HospDtls). The variable names correspond to those on the first two pages of the annotated Local Adjudication Report and on the annotated Discharge Abstract Form (see annotated forms).

3. Dataset structure and contents

The HospDtls file contains a single observation per participant-hospitalization. When a participant was hospitalized at one hospital at least overnight and then transferred to another hospital, there may be two records for that hospitalization event (use EVREF as the unique identifier of a single hospitalization event).

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
EVREF	Event Form Reference Number for that event

4. Condition of data

a. Known data errors: None at this time. The data have been edited.

b. Strengths and weaknesses of dataset items: This dataset is intended primarily as a way for investigators to obtain further details about particular hospitalizations that may be lost in the one-record-per-participant format of HOSP.sas7bdat. In addition, it is useful for investigators who may want to search out a subgroup based on ICD-9 discharge or diagnosis codes. There are several caveats to the use of this dataset however:

- This is not a good dataset to use to determine the number of hospitalizations per participant or even per type of diagnosis (e.g., cancer). In particular, as mentioned above, there may be two records for a single hospitalization event if the participant was transferred from one hospital to another during the same event.
- ICD-9 codes for diagnoses and procedures were obtained whenever possible, but for many records the codes were not readily available. Because ICD-9 codes are driven by reimbursement factors and are not necessarily consistent with clinical course during the actual hospitalization, investigators are encouraged to rely on the adjudicated diagnoses from the local adjudication form for overnight hospitalizations. ICD-9 codes for any outpatient cancers and fractures were not obtained, and investigators should refer to the cancer and fracture forms to identify the anatomical site involved
- There was considerable variation in how the ICD-9 codes were transcribed, with or without hyphens, with extra spaces, with or without leading or trailing zeroes, etc. The non-numeric characters have been removed and the spaces closed up so that there will be fewer variations in how the same code occurs in the dataset. These codes are also virtually impossible to edit, so there are many that may be non-valid codes.

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied

5. Dataset index formulation and key variable mapping

The HospDtls file is sorted by HABCID and then EVREF. The combination of these two variables is a unique identifier for each participant-hospitalization event. Note that it is possible for an EVREF to be duplicated between participants (one at Memphis and one at Pittsburgh), so it is very important to use the *combination* of EVREF and HABCID as the unique identifier.

6. General strategies for manipulating and merging the data

As noted above, there may be up to two records with the same combination of HABCID and EVREF, if a hospital transfer occurred. Analysts using this dataset should beware of the problems this could cause when merging this dataset with others.

DISEASE/CONDITION SPECIFIC DATASETS

1. General description

Each disease- or condition-specific dataset is derived from selected questions on the Local Adjudication Report or an adjudication report specific to that condition. Events are included whether they were the primary reason for hospitalization or occurred as a concurrent event during a hospitalization for another cause. Only events that were adjudicated as confirmed or probable are included. It is important to note that, except for outpatient revascularizations, some fractures, and some cancers, events that do not result in hospitalization are not captured and are therefore not included in this dataset. Analysts should be aware that disease- or condition-specific events that are not the primary reason for hospitalization represent a biased sample of non-primary events of that type. Note also that Cancer and Cardiovascular Disease Adjudication Forms are only completed for new events or when a previously uncertain event is confirmed. For example, if the first cancer event was a new diagnosis of breast cancer, a Cancer Adjudication Form would be completed for that event, but not for subsequent events involving the same breast cancer. However, if the participant subsequently is diagnosed with colon cancer as well, then a second Cancer Adjudication Form is completed for the event at which the colon cancer was diagnosed. If the first event is an uncertain diagnosis of breast cancer and a later event confirms the breast cancer, then there are two Cancer Adjudication Forms for the breast cancer. Similarly if an event is the participant's first diagnosis of angina, then a Cardiovascular Adjudication Form is completed. Subsequent angina attacks would not have an associated Cardiovascular Adjudication Form unless there was a concurrent new CVD diagnosis (e.g. MI or CHF). If the participant subsequently has an MI, then another Cardiovascular Adjudication Form is completed for that event. Again, if the first event of a particular type is uncertain and is followed by a confirmed diagnosis of that type, then there would be two Cardiovascular Adjudication Forms for that diagnosis. In addition, if any of the CVD diagnoses occurs during the hospitalization that results in the participant's death, a Cardiovascular Adjudication Form is completed, even if a confirmed diagnosis of that type has been documented before.

Each local adjudication form was completed to reflect current conditions resulting in or developing during the hospitalization. It does not reflect any medical history or chronic conditions unless they were acutely impacting the reason for the admission. Therefore, all co-morbidities are not listed on the local adjudication form unless directly related to the hospitalization.

In order to keep events data collection within a manageable volume, the D&D committee chose to restrict outpatient investigation to fractures, newly diagnosed cancers, and revascularizations. Data on other outpatient diagnoses and treatment were not collected with the justification that conditions requiring overnight hospitalizations would indicate a threshold of severity more likely to impact the primary outcome measure of function. Investigators should recognize that a sampling bias may exist for disease- or condition-specific diagnoses when both primary and concurrent diagnoses are included in analyses. A secondary condition may be captured in a medical record as a co-morbidity during an admission where the primary reason for hospitalization is a different condition. This secondary condition would not have been captured on other participants because they may not have any other hospitalizations in which the diagnosis was documented but may, in fact, have this condition under active treatment as an outpatient. It does not represent the prevalence of that secondary condition in the cohort.

Although data were collected for revascularizations, these were not adjudicated until recently. For new revascularizations that have occurred since adjudication began, there is now a new “other diagnosis” category. This new “other diagnosis” code is 15 and includes both CABG and PTCA. See “Other diagnosis dataset” below for further details about this category.

Each disease- or condition-specific dataset includes the variable, NUMxxx, which is the total number of hospitalization events in Health ABC involving a diagnosis of the type included in this specific dataset¹.

Please link to the Health ABC participant history (PH) for important per-participant level variables like gender, race, etc.

For each event captured involving a diagnosis of the specific type for that participant the following per-event variables are included (Xx indicates the xth event involving a diagnosis of type X; X varies by dataset²):

XxEVREF	Event Form Reference Number for xth event of the specified type
XxADMDT	Admission date for xth event of the specified type
XxDISDT	Discharge date for xth event of the specified type
XxLOSTA	Length of hospital stay for xth event of the specified type
XxCNCUR	Number of concurrent diagnoses during xth hospitalization event involving a diagnosis of the specified type (includes both “yes” and “possible” diagnoses, see page 2 of annotated Local Adjudication Report)
XxDS2EV	Number of days between enrollment (Baseline <u>clinic</u> visit) and xth event of the specified type
XxDXPRM	Primary diagnosis for that event (may be the same or different diagnosis from the event in question)

In addition, specific variables derived either from the Local Adjudication Report or from an adjudication report specific to the diagnosis are included for each event occurrence (again designated by the prefix Xx). Specifics regarding these variables are listed under Dataset Structure below.

¹ xxx=CANC for cancer; FRX for fractures; MI for myocardial infarctions; CHF for congestive heart failure; ANG for angina; HOSP for hospitalizations; BPH for benign prostatic hypertrophy; CAD for carotid artery disease; DIAB for diabetes; UPPER for upper GI; LOWER for lower GI; HERN for hernia; GALL for gall bladder disease; OASRG for osteoarthritis surgery; PAD for peripheral arterial disease; DEP for depression; DEM for dementia; PNEUM for pneumonia; CLD for chronic lung disease; STROK for stroke; TIA for transient ischemic attack; INF for infections; NEOP for neoplasms (benign); ENMI for endocrine, nutritional, metabolic or immune disorders; BLD for diseases of the blood and blood forming organs; MENT for mental disorders (not dementia or depression), NERVE for diseases of the nervous system and sense organs; CIRC for diseases of the circulatory system (excluding MI, angina, CHF); RESP for diseases of the respiratory system (excluding those listed individually above); DIGES for diseases of the digestive system (excluding those listed individually above); GUSYS for diseases of the genitourinary system (excluding BPH); SKIN for diseases of the skin; MSCT for diseases of the musculoskeletal system and connective tissue (excluding osteoarthritis); SxILL for symptoms, signs, and ill-defined conditions; INPOI for injuries and poisoning; and REVAS for revascularizations.

² BPH:Bx; Cancer: Cx; CVD:Vx; Carotid artery disease:Ax; Diabetes: Dx; Fracture: Fx; GI:Gx; Osteoarthritis:Jx; Peripheral arterial disease:Px; Psych:Sx; Pulmonary:Lx; Stroke and TIA:Tx; Other diagnoses:Ox

2. Cross reference of dataset names with exact source

A complete list of variable names can be found in Appendix V: Proc Contents (search under dataset name). Variable names can also be found on the annotated forms.

3. Dataset structure and contents

Each disease- or condition-specific file contains a single observation per participant and all participants enrolled are represented (n=3075), regardless of whether they have experienced an event of that type. Those who have experienced an event of the specific type can be pulled out using NUMxxx>0.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
HxEVREF	Event Form Reference Number for the xth event; this number can be used to identify this event within the hospitalization datasets, particularly the Hospitalization Details dataset (HospDtls.sas7bdat)

a. BPH dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each BPH event :

BxADJUD	Adjudication indicator variable (0=no event; 1=definite; 2=probable)
BxDIAG	Indicator variable indicating whether the xth BPH event was the primary reason for hospitalization or a concurrent diagnosis (0=no hospitalized BPH event, 1=BPH as primary reason for hospitalization; 2=BPH as concurrent event)

Plus the following variables from Question 15, page 9 of the Local Adjudication (at least one of these criteria must be present for a positive adjudication):

BxBPH1	Current or prior history of symptoms of urinary obstruction
BxBPH2	Digital exam evidence of prostate enlargement
BxBPH3	Treatment with specific medications
BxBPH4	Surgical procedure (current event or history)
BxBPHDT	Date of surgical procedure
BxBPH5	(Urinary retention documented by IVP, ultrasound, catheterization)

b. Cancer dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each cancer event:

CxADJUD	Adjudication indicator variable (0=no event; 1=definite; 2=probable)
CxDIAG	Indicator variable indicating whether the xth cancer event was an out-patient event, the primary reason for hospitalization, or a concurrent diagnosis (0=no cancer event, 1=cancer as primary reason for hospitalization; 2=cancer as concurrent diagnosis; 3=outpatient cancer event)

There are also 5 variables for each cancer event corresponding to the variables from the Cancer Adjudication Report (see annotated forms).

c. Cardiovascular dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each cardiovascular event.

VxMI	Indicator variable indicating whether the xth cardiovascular event included an MI (0=no hospitalized MI event, 1=MI as primary reason for hospitalization, 2=MI as concurrent diagnosis)
VxMIAD	Adjudication indicator variable for MI (0=no event; 1=definite; 2=probable)
VxANGIN	Indicator variable indicating whether the xth cardiovascular event included angina (0=no angina, 1=angina as primary reason for hospitalization, 2=angina as concurrent diagnosis; 3=outpatient angina event)
VxANGAD	Adjudication indicator variable for angina (0=no event; 1=definite; 2=probable)
VxCHF	Indicator variable indicating whether the xth cardiovascular event included CHF (0=no hospitalized CHF event, 1=CHF as primary reason for hospitalization, 2=CHF as concurrent diagnosis)
VxCHFAD	Adjudication indicator variable for CHF (0=no event; 1=definite; 2=probable)

There are also 36 variables for each CVD event corresponding to the variables from the Cardiovascular Disease Adjudication Report (see annotated forms).

A Cardiovascular Adjudication Report is completed for incident myocardial infarction, angina and/or congestive heart failure. If the participant meets the criteria for MI, then angina is not coded as a co-morbid condition on the Local Adjudication Report unless there is at least one episode of post infarction angina or the participant is admitted for angina and rules out for MI but subsequently has an MI (e.g., post-CABG MI).

The Cardiovascular Adjudication Report is designed to identify clinical diagnoses. Cases with cardiovascular procedures such as bypass surgery cannot be identified using this dataset, because only procedures occurring during an incident angina event are recorded.

d. Carotid Artery Disease dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each carotid artery disease event:

AxADJUD	Adjudication indicator variable (0=no event; 1=definite; 2=probable)
AxDIAG	Indicator variable indicating whether the xth carotid artery disease event was the primary reason for hospitalization or a concurrent diagnosis (0=no hospitalized carotid artery disease event, 1=carotid artery disease as primary reason for hospitalization; 2=carotid artery disease as concurrent diagnosis)

Plus the following variables are included from Question 6, page 4 of the Local Adjudication Report (at least one of these criteria must be present for a positive adjudication):

AxCAD1	Symptomatic disease/carotid artery disease listed on discharge summary
AxCAD2	Symptomatic disease/abnormal findings on angiogram or Doppler flow
AxCAD3	Vascular or surgical procedure to improve flow to ipsilateral brain

e. Death Dataset: The Death dataset is derived from questions on page 2 of the Local Adjudication Report, the Report of Death, the Final Death Adjudication Report, and the Decedent Proxy Interview (see annotated forms). Please link to the Health ABC participant history (PH) file for important per-

participant generic variables including: VStatus (1=Alive, 2=Dead, .Z=missed last visit), DTLASTCT (Date of last Health ABC contact and DOD (Date of Death from Report of Death, or Event Form if no Report of Death).

The Death dataset also includes the following variables, which, of course, are per-participant rather than per-event in the case of deaths:

EVREF	Event Form Reference Number for the death
WFADMDT	Admission date for the hospitalization leading to death (if applicable)
WFDXPRM	Primary diagnosis for hospitalization leading to death (if applicable)
LOSTA	Length of stay in the hospitalization before death (if applicable)
INHOSP	Where death occurred (0=not in hospital; 1=in hospital)
CNCUR	Number of concurrent diagnoses during the hospitalization leading to death (if applicable)
DS2DTH	Number of days between enrollment (Baseline <u>clinic</u> visit) and xth event of the specified type

See also the annotated Report of Death, Final Death Adjudication Report, and Decedent Proxy Interview for further variables included.

Although adjudication of deaths can never be up-to-the-minute, the VStatus variable is as close to up-to-the-minute as can be obtained. As soon as a clinic receives notification of a participant's death, completes an Event Form for that death, and enters it, the VStatus variable will reflect the death, even though adjudication may not be complete for several months. (VStatus is set to .Z when the participant has no data for the last visit s/he was scheduled for and the window is closed.) The date of death (DOD) is taken from the death certificate (as entered on the Report of Death if completed), or from the Event Form (if there is no Report of Death yet). Thus, some dates of death are only as reported by the person informing the clinic of the death, not from the death certificate.

The variable INHOSP has to takes advantage of question #8 on the Report of Death (adjudicated location of death). If no Report of Death has been entered for a known death (VSTATUS=2 or non-missing DOD), then INHOSP is missing. Thus a non-missing value of INHOSP is always adjudicated location of death information.

f. Diabetes dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each diabetes event:

DxADJUD	Adjudication indicator variable (0=no event; 1=definite; 2=probable)
DxDIAG	Indicator variable indicating whether the xth diabetes event was the primary reason for hospitalization or a concurrent diagnosis (0=no hospitalized diabetes event; 1=diabetes as primary reason for hospitalization; 2=diabetes as concurrent diagnosis)

Plus the following variables from Question 16, page 9 of the Local Adjudication Report (at least one of these criteria must be present for a positive adjudication):

DxDIAB1	Hypoglycemia
DxDIAB2	Hyperosmolar coma

Diabetes or hyperglycemia, as either a primary reason for hospitalization or as a concurrent diagnosis during hospitalization, not meeting the above criteria is coded under "Other diagnosis" code "3".

g. Fracture dataset: For the Fracture dataset, the fracture SOP required completion of a separate Fracture Adjudication Report for each fracture, even if more than one fracture occurred during the same Health ABC event. Concurrent fractures can be recognized by identical fracture dates (FxFRXDT). In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each fracture:

FxADJUD	Adjudication indicator variable (0=no event; 1=definite; 2=probable)
FxDIAG	Indicator variable indicating whether the xth fracture event was an outpatient event, the primary reason for hospitalization, or a concurrent diagnosis (0=no fracture event; 1=fracture as primary reason for hospitalization; 2=fracture as concurrent diagnosis; 3=outpatient fracture)

There are also 6 variables for each fracture event corresponding to the variables from the Fracture Adjudication Report (see annotated forms).

h. GI dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each GI event:

GxUPPER	Indicator variable indicating whether the xth GI event included a UGI diagnosis (0=no hospitalized UGI diagnosis, 1=UGI diagnosis as primary reason for hospitalization, 2=UGI diagnosis as concurrent diagnosis)
GxUPPAD	Adjudication indicator variable for upper GI event (0=no event; 1=definite; 2=probable)
GxLOWER	Indicator variable indicating whether the xth GI event included a LGI diagnosis (0=no hospitalized LGI diagnosis, 1=LGI diagnosis as primary reason for hospitalization, 2=LGI diagnosis as concurrent diagnosis)
GxLOWAD	Adjudication indicator variable for lower GI event (0=no event; 1=definite; 2=probable)
GxHERN	Indicator variable indicating whether the xth GI event included a hernia diagnosis (0=no hospitalized hernia diagnosis, 1=hernia diagnosis as primary reason for hospitalization, 2=hernia diagnosis as concurrent diagnosis)
GxHRNAD	Adjudication indicator variable for hernia event (0=no event; 1=definite; 2=probable)
GxGALL	Indicator variable indicating whether the xth GI event included gallbladder disease (0=no hospitalized gallbladder disease, 1=gallbladder disease as primary reason for hospitalization, 2=gallbladder disease as concurrent diagnosis)
GxGALAD	Adjudication indicator variable for gall bladder event (0=no event; 1=definite; 2=probable)

The following variables from Questions 12-14 & 17, pages 7-9 of the Local Adjudication Report are also included for each GI event (at least one of the criteria for each diagnosis (asterisked) must be present for a positive adjudication):

GxUGI1	Duodenal ulcer
GxUGI2	Gastric ulcer
GxUGI3	Ulcer, site unspecified
GxUGI4	Gastritis/gastric erosions/duodenal erosions
GxUGI5	Esophageal/gastric varices
GxUGI6	Other esophageal
GxUGI7	Neoplasia
GxUGI8	Other UGI diagnosis
GxUGIOT	"Other" UGI diagnosis text
GxUGI9	No source found
GxUGBLD	Did UGI condition result in bleeding? (0=No; 1=Yes; 9=Unknown)
GxUGCR1*	Upper endoscopy diagnosis
GxUGCR2*	Transfusion
GxUGCR3*	Upper GI series diagnosis
GxUGCR4*	Nasogastric tube
GxLGI1	Diverticulosis/diverticulitis
GxLGI2	Angiodysplasia
GxLGI3	Neoplasia
GxLGI4	Colitis
GxLGC01	Radiation colitis
GxLGC02	Ischemic colitis
GxLGC03	Infectious colitis
GxLGC04	Ulcerative colitis
GxLGI5	Hemorrhoids
GxLGI6	"Other" LGI diagnosis
GxLGIOT	"Other" LGI diagnosis text
GxLGI7	No LGI source found
GxLGBLD	Did the LGI condition result in bleeding? (0=No; 1=Yes; 9=Unknown)
GxLGCR1*	Lower endoscopy diagnosis
GxLGCR2*	Transfusion
GxLGCR3*	Angiography diagnosis
GxLGCR4*	Barium enema/lower GI series diagnosis
GxHERN1	Inguinal hernia
GxHERN2	Incisional hernia
GxHERN3	Umbilical hernia
GxHRNC1*	Noted on admission physical exam
GxHRNC2*	Hernia repair
GxGALL1	Gallstones (cholelithiasis)
GxGALL2	Acute cholecystitis
GxGALL3	Common bile duct stone (choledocholithiasis)
GxGALC1	Diagnosis on ultrasound of bladder/common bile duct ERCP/CT scan/radionuclide scan
GxGALC2	Cholecystectomy

i. Osteoarthritis / Joint Surgery dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each osteoarthritis surgery event:

JxADJUD	Adjudication indicator variable (0=no event; 1=definite)
---------	--

JxDIAG Indicator variable indicating whether the xth OA event was the primary reason for hospitalization or a concurrent diagnosis (0=no hospitalized OA/joint surgery event; 1=OA/joint surgery as primary reason for hospitalization; 2=OA/joint surgery as concurrent diagnosis)

Plus the following variables from Question 19, page10 of the Local Adjudication Report (see annotated forms):

JxOAKNE Knee joint involved
JxOAHIP Hip joint involved
JxOAOTH Other joint involved
JxOAOT1 Which other joint involved (text variable)
JxOAKLR Which knee involved (1=right; 2=left; 3=both)
JxOAHLR Which hip involved (1=right; 2=left; 3=both)
JxOADX Diagnosis for operated joint (1=includes radiology report; 2=no radiology report)
JxOADX1 Other diagnosis: Rheumatoid arthritis
JxOADX2 Other diagnosis: Aseptic necrosis
JxOADX3 Other diagnosis: Congenital hip disease
JxOADX4 Other diagnosis: Other
JxOADXO "Other" diagnosis type (text variable)

j. Peripheral arterial disease dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each PAD event:

PxADJUD Adjudication indicator variable (0=no event; 1=definite; 2=probable)
PxDIAG Indicator variable indicating whether the xth PAD event was the primary reason for hospitalization or a concurrent diagnosis (0=no hospitalized PAD event; 1=PAD as primary reason for hospitalization; 2=PAD as concurrent diagnosis)

Plus the following variables from Question 7, page 4 of the Local Adjudication (at least one of these criteria must be present for a positive adjudication):

PxPADDX PAD Diagnosis
PxPAD1 Blockage or ulcerated plaque on ultrasound/angiogram
PxPAD2 Absence of pulse on Doppler / all major lower extremity vessels
PxPAD3 Positive exercise test for lower extremity claudication
PxPAD4 Surgery/angioplasty/thrombolysis for PAD
PxPAD5 Amputation of ≥ 1 toes or part of lower extremity due to ischemia/gangrene
PxPAD6 Exertional pain relieved by stress and diagnosed claudication or ankle-arm systolic GP ration<0.8
PxPAD7 Ultrasound/angiography/CT-demonstrated diagnosis, other than lower extremities
PxPAD8 Surgical or vascular procedure other than lower extremities

k. Psychiatric event dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each psychiatric event:

SxDEM	Indicator variable indicating whether the xth psychiatric event included a diagnosis of dementia (0=no hospitalized psychiatric event; 1=dementia as primary reason for hospitalization; 2=dementia as concurrent diagnosis)
SxDEMAD	Adjudication indicator variable for dementia (0=no event; 1=definite; 2=probable)
SxDEP	Indicator variable indicating whether the xth psychiatric event included a diagnosis of depression (0=no hospitalized psychiatric event; 1=depression as primary reason for hospitalization; 2=depression as concurrent diagnosis)
SxDEPAD	Adjudication indicator variable for depression (0=no event; 1=definite; 2=probable)

Plus the following variables from Question 18, page 10 of the Local Adjudication Report (see annotated forms):

SxDEM1	Chronic, progressive cognitive decline not explained by other disease process
SxDEM2	In presence of disease process affecting cognition, a history of dementia before onset or after treatment/resolution of other disease process
SxDEM3	Use of Rx medication specific to Alzheimer's disease

There are no specific criteria used for adjudication of depression. In addition, information on psychiatric hospitalizations is not routinely collected. Therefore, there are few cases of depression as the primary reason for hospitalization.

l. Pulmonary event dataset: In addition to the generic variables in the Health ABC participant history (PH) file, the following are included for each pulmonary event:

LxPNEUM	Indicator variable indicating whether the xth pulmonary event included a pneumonia diagnosis (0=no hospitalized pneumonia event, 1=pneumonia diagnosis as primary reason for hospitalization, 2=pneumonia diagnosis as concurrent diagnosis)
LxPNMAD	Adjudication indicator variable for pneumonia (0=no event; 1=definite; 2=probable)
LxCLDAD	Adjudication indicator variable for a chronic lung disease event (0=no event; 1=definite; 2=probable)
LxCLD	Indicator variable indicating whether the xth pulmonary event included a chronic lung disease diagnosis (asthma, bronchitis, COPD, or emphysema) (0=no hospitalized CLD diagnosis, 1=CLD diagnosis as primary reason for hospitalization, 2=CLD diagnosis as concurrent diagnosis)

Plus the following variables from Questions 10 & 11, page 6 of the Local Adjudication Report (see annotated forms):

LxASTHM	Asthma diagnosis
LxBRONC	Bronchitis diagnosis
LxCOPD	COPD diagnosis
LxEMPHY	Emphysema diagnosis
LxCOPD1	Chest x-ray consistent with COPD

LxCOPD2	Spirometry results consistent with COPD
LxCOPD3	Arterial blood gases consistent with respiratory insufficiency
LxCOPD4	Physical exam consistent with COPD
LxCOPD5	Treatment with bronchodilators, corticosteroids, or oxygen
LxPNEU1	Symptoms of cough/fever/sputum production, rales, or dullness to percussion
LxPNEU2	Chest x-ray showing infiltrate, consolidation, cavitation, or pleural effusion
LxPNORG	Organism identified (0=No; 1=Yes; 9=Unknown)
LxPNORS	Specific organism (text variable)

m. Stroke/TIA event dataset: In addition to the generic variables in the Health ABC participant history (PH) file, are included for each stroke/TIA event :

TxSTROK	Indicator variable indicating whether the xth stroke/TIA event included a stroke diagnosis (0=no hospitalized stroke diagnosis, 1=stroke diagnosis as primary reason for hospitalization, 2=stroke diagnosis as concurrent diagnosis)
TxSTRAD	Adjudication indicator variable for stroke (0=no event; 1=definite; 2=probable)
TxTIA	Indicator variable indicating whether the xth stroke/TIA event included a TIA diagnosis (0=no hospitalized TIA diagnosis, 1=TIA diagnosis as primary reason for hospitalization, 2=TIA diagnosis as concurrent diagnosis)
TxTIAAD	Adjudication indicator variable for TIA (0=no event; 1=definite; 2=probable)
TxCVA	Stroke subtype (1=hemorrhagic; 2=ischemic; 3=unknown)

Plus the following variables from Questions 8 & 9, page 5 of the Local Adjudication Report (all of the asterisked criteria must be checked for a positive adjudication):

TxCVHEM	Hemorrhagic stroke location (1=subarachnoid; 2=intracerebral; 9=unknown; 8=N/A)
TxCVISC	Ischemic stroke type (1=lacunae; 2=embolic; 3=atherosclerotic; 9=unknown)
TxCVA1*	Rapid onset of neurologic deficit attributed to obstruction/rupture of arterial system
TxCVA2*	Deficit lasting >24 hours (unless death)
TxCVA3*	No evidence of tumor, trauma, infection, or other non-ischemic cause
TxCVA4	New CT/MRI lesion consistent with clinical presentation
TxCVPRO	Was stroke procedure related? (0=No, 1=Yes, 9=unknown)
TxTIA1*	≥1 episode of focal neurologic deficit lasting 30 sec-24 hours
TxTIA2*	Maximal deficit in < 5 minutes
TxTIA3*	Complete resolution
TxTIA4*	No head trauma before onset
TxTIA6*	No evidence of seizure

n. "Other" diagnosis dataset: The Local Adjudication Report also captures 15 possible "other" diagnoses, besides the 20 specific diagnoses listed on page 2 of the report. The variables OxDXPRM are formatted with the format PRMDXF, which matches these codes to the variable values. The analyst should be sure to download the latest Formats catalog to ensure that these variables are formatted correctly. These "other" diagnoses include:

- 1 = Infectious diseases
- 2 = Neoplasms (benign)

- 3 = Endocrine, nutritional, metabolic diseases, immune disorders
- 4 = Diseases of the blood & blood forming organs
- 5 = Mental disorders (not dementia & depression)
- 6 = Diseases of nervous system & sense organs
- 7 = Diseases of the circulatory system (excluding MI, angina, CHF etc)
- 8 = Diseases of the respiratory system (excluding COPD etc)
- 9 = Diseases of the digestive system (excluding upper GI disease etc)
- 10 = Diseases of the genitourinary system (excluding BPH)
- 11 = Diseases of the skin
- 12 = Diseases of the musculoskeletal system & connective tissue (excluding osteoarthritis)
- 13 = Symptoms, signs & ill defined conditions
- 14 = Injury & poisoning (excluding fractures)
- 15 = Revascularizations (CABG and PCTA)³

Similar to the other disease-specific datasets, each diagnosis captured with an "other" diagnosis code for a particular participant is documented by the following per-diagnosis variables, in addition to those variables in the Health ABC participant history (PH) file, (Ox indicates the xth "other" diagnosis; note that one hospitalization event may be represented by up to three "other" diagnoses):

OxINF	Indicator variable indicating whether the xth "other" event included a infectious disease diagnosis (0=no hospitalized infectious disease diagnosis, 1=infectious disease as primary reason for hospitalization, 2=infectious disease as concurrent diagnosis)
OxINFAD	Adjudication indicator variable for infection (0=no event; 1=definite; 2=probable)
OxNEOP	Indicator variable indicating whether the xth "other" event included a neoplasm diagnosis (0=no hospitalized neoplasm diagnosis, 1=neoplasm as primary reason for hospitalization, 2=neoplasm as concurrent diagnosis)
OxNEOAD	Adjudication indicator variable for neoplasm (0=no event; 1=definite; 2=probable)
OxENMI	Indicator variable indicating whether the xth "other" event included a endocrine/nutritional/metabolic/immune disorder diagnosis (0=no hospitalized endocrine/nutritional/metabolic/immune disorder diagnosis, 1=endocrine/nutritional/metabolic/immune disorder as primary reason for hospitalization, 2=endocrine/nutritional/metabolic/immune disorder as concurrent diagnosis)
OxENMAD	Adjudication indicator variable for endocrine, nutritional, metabolic or immune disorder (0=no event; 1=definite; 2=probable)
OxBLD	Indicator variable indicating whether the xth "other" event included a blood disorder diagnosis (0=no hospitalized blood disorder diagnosis, 1=blood disorder as primary reason for hospitalization, 2=blood disorder as concurrent diagnosis)
OxBLDAD	Adjudication indicator variable for disease of blood and blood forming organs (0=no event; 1=definite; 2=probable)

³ Adjudication of revascularizations is now complete. After the decision was made at the February 2003 Steering Committee meeting to go back and recode these, they were identified by looking for all Local Adjudication Reports with a cardiovascular diagnosis listed as "other," and then reviewing each of these cases to determine whether a revascularization had been done. Note that this does NOT include outpatient revascularizations.

OxMENT	Indicator variable indicating whether the xth "other" event included a mental diagnosis (0=no hospitalized mental diagnosis, 1=mental diagnosis as primary reason for hospitalization, 2=mental diagnosis as concurrent diagnosis)
OxMNTAD	Adjudication indicator variable for mental disorder other than dementia or depression (0=no event; 1=definite; 2=probable)
OxNERVE	Indicator variable indicating whether the xth "other" event included a nervous system/sense organ disorder diagnosis (0=no hospitalized nervous system/sense organ disorder diagnosis, 1=nervous system/sense organ disorder as primary reason for hospitalization, 2=nervous system/sense organ disorder as concurrent diagnosis)
OxNRVAD	Adjudication indicator variable for disease of nervous system and sense organs (0=no event; 1=definite; 2=probable)
OxCIRC	Indicator variable indicating whether the xth "other" event included a circulatory disorder diagnosis (0=no hospitalized circulatory disorder diagnosis, 1=circulatory disorder as primary reason for hospitalization, 2=circulatory disorder as concurrent diagnosis)
OxCRCAD	Adjudication indicator variable for disease of the circulatory system other than those in the Cardiovascular dataset (0=no event; 1=definite; 2=probable)
OxRESP	Indicator variable indicating whether the xth "other" event included a respiratory disorder diagnosis (0=no hospitalized respiratory disorder diagnosis, 1=respiratory disorder as primary reason for hospitalization, 2=respiratory disorder as concurrent diagnosis)
OxRSPAD	Adjudication indicator variable for respiratory disease other than those in the Pulmonary dataset (0=no event; 1=definite; 2=probable)
OxDIGES	Indicator variable indicating whether the xth "other" event included a digestive disorder diagnosis (0=no hospitalized digestive disorder diagnosis, 1=digestive disorder as primary reason for hospitalization, 2=digestive disorder as concurrent diagnosis)
OxDGSAD	Adjudication indicator variable for disease of the digestive system other than those in the GI dataset (0=no event; 1=definite; 2=probable)
OxGUSYS	Indicator variable indicating whether the xth "other" event included a genitourinary system diagnosis (0=no hospitalized genitourinary system diagnosis, 1=genitourinary system diagnosis as primary reason for hospitalization, 2=genitourinary system diagnosis as concurrent diagnosis)
OxGUAD	Adjudication indicator variable for genitourinary disease other than BPH (0=no event; 1=definite; 2=probable)
OxSKIN	Indicator variable indicating whether the xth "other" event included a skin disease diagnosis (0=no hospitalized skin disease diagnosis, 1=skin disease as primary reason for hospitalization, 2=skin disease as concurrent diagnosis)
OxSKNAD	Adjudication indicator variable for skin disease (0=no event; 1=definite; 2=probable)
OxMSCT	Indicator variable indicating whether the xth "other" event included a musculoskeletal/connective tissue disorder diagnosis (0=no hospitalized musculoskeletal/connective tissue disorder diagnosis, 1=musculoskeletal/connective tissue disorder as primary reason for hospitalization, 2=musculoskeletal/connective tissue disorder as concurrent diagnosis)

OxMCTAD	Adjudication indicator variable for disease of the musculoskeletal system and connective tissue other than osteoarthritis (0=no event; 1=definite; 2=probable)
OxSxILL	Indicator variable indicating whether the xth "other" event included a symptom, sign, or ill-defined condition (0=no hospitalized symptom, sign, or ill-defined condition, 1=symptom, sign, or ill-defined condition as primary reason for hospitalization, 2=symptom, sign, or ill-defined condition as concurrent diagnosis)
OxILLAD	Adjudication indicator variable for symptom, sign, or other ill-defined condition (0=no event; 1=definite; 2=probable)
OxInPoi	Indicator variable indicating whether the xth "other" event included a injury or poisoning (0=no hospitalized injury or poisoning, 1=injury or poisoning as diagnosis)
OxINPAD	Adjudication indicator variable for injury or poisoning (0=no event; 1=definite; 2=probable)
OxREVAS	Indicator variable indicating whether the xth "other" event included a revascularization (0=no revascularization, 1=revascularization as diagnosis)
OxRVAD	Adjudication indicator variable for revascularization (0=no event; 1=definite; 2=probable)

4. Condition of data

a. Known data errors: None at this time. The data have been edited.

b. Strengths and weaknesses of dataset items: It is again important to note that, except for revascularizations, some fractures, and some cancers, events that do not result in hospitalization are not captured and are therefore not included in these datasets. Analysts should be aware that events that are not the primary reason for *hospitalization* represent a biased sample of non-primary events of the specified type. To restrict analysis to primary events of the specified type, only events with XxDIAG=1 should be used (or, when a dataset contains more than one possible diagnosis, e.g., Cardiovascular, the individual indicator variables, in this case VxMI, VxANGIN, and VxCHF, would be restricted to 1).

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied

5. Dataset index formulation and key variable mapping

The disease- or condition-specific files are sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

SPECIAL MISSING VALUE CODES

SAS allows for stratification of missing values. The following missing values have been assigned:

. = 'Missing Form'
.A = 'A:Not Applicable'
.M = 'M:Missing'
.N = 'N:Not Required'

Description

.: Missing Form

Used when a value is missing because the entire form has not been entered.

A: Not Applicable

Used when a value is missing but the value is not required (due to simple skip pattern logic)

M:Missing

Used to flag missing values when the value is required (i.e., true missing values).

N:Not Required

Used when a value is missing but the value is not required (not due to simple skip pattern logic). For example, for checkbox variables which are "Check all that apply" each one, individually, is not required. In these cases, a summary calculated variable (not included in the dataset) was used to edit missing responses. Some variables whose skip pattern logic is non-standard (i.e., the skip pattern involves several variables and forms) also have .N flags when missing, whether or not a response was required due to the skip pattern.

General Strategies for Using Special Missing Values

In SAS, when using special missing values in logical expressions, the missing value is no longer only equal to '.' To express a value equal to missing, the code should be written: `<= .z` or alternately: `le .z`

To express a value not equal to missing, the code should be written `>.z` or alternately: `gt .z`. `.Z` is the greatest value of missing available in SAS.